PREDICTION OF HARMFUL ALGAE BLOOMS IN SABAH USING DEEP LEARNING MODEL

MOHD FIRDAUS BIN PATITINGI

FACULTY OF COMPUTING AND INFORMATICS

UNIVERSITI MALAYSIA SABAH

2022

PREDICTION OF HARMFUL ALGAE BLOOMS IN SABAH USING DEEP LEARNING MODEL

MOHD FIRDAUS BIN PATITINGI

THESIS SUBMITTED IN PARTIAL FULFILLMENT FOR THE DEGREE OF BACHELOR OF COMPUTER SCIENCE WITH HONOURS (NETWORK ENGINEERING)

FACULTY OF COMPUTING AND INFORMATICS
UNIVERSITI MALAYSIA SABAH

2022

NAME	**:** MOHD FIRDAUS BIN PATITINGI
MATRIC NUMBER	: BI18110040
TITTLE	: PREDICTION OF HARMFUL ALGAE BLOOMS IN SABAH USING DEEP LEARNING MODEL
BACHELOR	: BACHELOR OF COMPUTER SCIENCE (NETWORK ENGINEERING)
DATE	: 16 FEBRUARY 2022

VERIFIED BY:

1. SUPERVISOR	**SIGNATURE**
MADAM SALMAH BINTI FATTAH	SALMAH FATTAH Lecturer Faculty of Computing and Informatics Universiti Malaysia Sabah

UMS
UNIVERSITI MALAYSIA SABAH

DECLARATION

I hereby declare that the material in this thesis is my own except for quotations, accepts equations summaries and reference, which have been duly acknowledged.

16 FEBRUARY 2022

MOHD FIRDAUS BIN PATITINGI

BI18110040

ACKNOWLEDGEMENT

The success and the final outcome of this project required guidance and assistance from different sources and I feel extremely fortunate to have got this all along the completion of my project. Whatever I have done is largely due to such guidance and assistance and I would not forget to thank them. I express our sincere thanks to my supervisor madam Salmah Binti Fattah for all the help and infrastructure provided to me to complete this project successfully and her valuable guidance. I also owe my profound gratitude to my project examiner professor Dr. Jason Teo Tze Wi and Dr. Florence Sia Fui Sze who took keen interest in my project and guided me all along during presentation till the completion of this project work by providing all the necessary information for developing a good system. I are thankful to and fortunate enough to get constant encouragement, support and guidance from all the Teaching staff of the Faculty of Computing and Informatics (FCI) UMS which helped me in successfully completing my final year project work.

MOHD FIRDAUS BIN PATITINGI

BI18110040

ABSTRACT

Harmful algal blooms (HAB) incident have been increasingly reported in the country and it became a critical global environmental concern which might put economic development and sustainability at risk. However, the analysis and accurate prediction of algae blooms remains a challenging scientific problem. In this project, a method based on deep learning is an approach to analysis and predict highly dynamic to the incident of HAB. People expect that such a system will significantly facilitate researchers, local administrators and civilians in monitoring water bodies and immediately solve any excessive algae growth. From the results of this study, it can be proven that the deep learning model make a better generalization and greater accuracy in predicting algae blooms than a traditional shallow neural network does.

ABSTRAK

Insiden 'algal bloom' telah menigkat di Negara ini dan ia menjadi kebimbangan besar kepada alam sekitar global yang boleh mengakibatkan pembangunan ekonomi dan kemampanan berisiko. Walau bagaimanapun, analisis dan ketepatan ramalan algal bloom tetap menjadi cabaran dari segi masalah saintifik. Dalam projek ini, kaedah 'deep learning' telah diterapkan untuk menganalisis dan meramalkan dengan sangat dinamik terhadap kejadian 'algal bloom'. Orang ramai akan menjangkakan bahawa sistem ini akan memudahkan penyelidik, pentadbir tempatan dan orang awam dalam memantau keadaan permukaan air dan segera dapat menyelesaikan sebarang pertumbuhan alga yang berlebihan. Daripada hasil kajian ini, dapat dibuktikan bahawa model 'deep learning' dapat membuat generalisasi yang lebih baik dan ketepatan yang lebih tinggi dalam meramalkan 'algal bloom' daripada apa yang model tradisional 'neural network' lakukan.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLE

CHAPTER 1

INTRODUCTION

1.1 OVERVIEW CHAPTER

This section will focus on the background of the project where the purpose is being highlighted with further elaboration on it. Besides, the problem statements, project objective and project scope of this project is analysed and take into consideration as well as the overview of the report throughout the process of developing this project. This section is crucial as it is the initial step that needs clear picture on the initiatives, analysation and requirements of the projects.

1.2 PROBLEM BACKGROUND

Harmful algae blooms monitoring in Malaysia is currently conducted by the Fisheries Research Institutes (FRI) and the Sabah Fisheries Department (SFD). Shellfish toxicity monitoring was confined to selected locations on the coast of Sabah due to its frequent outbreak of P. Bahamans blooms(Jipanin et al. 2019). However, sampling interval and frequency of monitoring exercises conducted are not sufficient to monitor the Harmful algae blooms occurrence. Most of the bloom onset and declining occurred within a relatively in short duration. Furthermore, monitoring program is slow down by lack of trained manpower and expertise in phytoplankton ecology which could lead to misidentification of harmful species(Teen, Gires, and Pin 2012). The recent activity in harmful algae blooms surveillance susceptibility tests across new broad areas require a significant new resource and yet there are limited in computational tools for the research.

1.3 PROBLEM STATEMENTS

The main problem to make this study is there were limited computational tools to detect the Harmful Algae Bloom in the seawater. It only have manually method by taking the sample and record the data. That method not sufficient to prevent the harmful algae blooms grow in the water. From that problem, there is less efficient method to perform future predictions of incident harmful algae bloom. For example, the prediction of algae bloom using Landsat thematic Mapper (TM) data only get 74% accuracy that mean the problem still cannot be solve. Other method was more to prevent the harmful algae bloom from occur but people don't know when the problem come again.

1.4 PROJECT OBJECTIVES

- To design a model which can detect the Harmful Algae Bloom using Deep Learning Model.
- To implement the deep learning model on harmful algae bloom prediction system using web based.
- To test and evaluate the effectiveness and system performance toward the harmful algae blooms in Sabah

1.5 PROJECT SCOPE

The scope of this project is defined that it will implemented on research place of Borneo Marine Research Institute (IPMB). The scope of this project will develop a system that can perform predictions of incident harmful algae bloom. The user of this project is employee of Borneo Marine Research Institute (IPMB) that design restrictions for harvesting more reliable and selective shellfish and minimising economic impacts on local communities while maintaining human health.

1.6 CONCLUSION

In conclusion, this chapter can help for better understanding about harmful algal bloom that occur around the world especially Malaysia. The problem that now currently having, the scope of this project and the most importantly is the objectives of this project make a clear vision how to get a better result in monitoring and prediction.

CHAPTER 2

LITERATURE REVIEW

2.1 OVERVIEW CHAPTER

This chapter will discuss the study that has been made with a review of previous related research. The discussion involves the study Similar Existing System and also the implication of literature review on the proposed system. A literature review is a study of scholarly references in a specific subject area and often knowledge within a time span in a specific subject area. An analysis of literature may be a straightforward overview of the sources typically has an organizational pattern and incorporate both summary and synthesis. A literature review also provides a solid background for a research paper's investigation. Writing a literature review involves finding relevant publications such as books and journal articles, critically analysing them and explaining what the outcome is. A good literature review doesn't just summarize sources, it analyses, synthesis, and critically evaluates to give a clear picture of the state of knowledge on the subject. Comprehensive knowledge of the literature of the field is essential to most research papers.

2.2 HARMFUL ALGAE BLOOMS

Various factors can cause harmful algal blooms, such as an increase in nutrients from an influx of anthropogenic contaminants produced by households, factories, farmland, or other sources. The rate of algae growth can increase rapidly and large algal colonies sometimes formed and covering many area of water. It releasing paralytic, neurotoxic, diarrheic, amnesic and azaspiracid toxins that causing shellfish poisoning in water and lead to harm the sea creatures, birds and even humans(Anderson 2009). It also when HAB decay, it consume large amount of oxygen and it causing anoxia and harmless to the sea creatures that can lead to dead.

The HAB also can be form by the changes in the chlorophyll concentration in water quantity. But from this incident it can be prevent by raising the water level could moderate the existing of HAB. The eutrophication stratum became vertically blended when increasing the water and it reduce the time for algae remained on the surface. In this way, algal propagation can be reduced because of dilution and dispersion of the nutrients(Anderson 2009).

In Malaysia, HABs events have been increasingly reported over the last decade. This may be the cause of increasing utilization of coastal inhabitants, aquaculture and other human activities that effect negatively to eutrophication at the waters. Industrialization and commercial agricultural plantation also led to nutrient decreasing and nutrient enrichment increase in the water and marine systems produce the algae-bloom (Teen et al. 2012).

This HAB make the responsible authorities needs to build a better HABs monitoring and prediction in the country. Current HABs studies in the country are insufficient to safeguard the public health and seafood safety. In Sabah, the authorities make sampling program for PSP toxins in shellfish, fish and plankton due to the continuously happen of HAB in the west coast of Sabah. The samples are collected biweekly in at least 12 districts of Sabah. A public warning is issued when toxin levels reach 80 µg/100g or the density of Pyrodinium cells reach 7,000 cells/liter. Results from more than 25 years of monitoring indicate that PSP levels vary widely for different locations, at different times of the year and for different HAB species (Anon n.d.).

2.3 DEEP LEARNING MONITORING.

It is important to constantly monitor water surface and identify any algae growing up so that an action to prevent it can be taken. All the approaches capture images in various orientations, making it very unlikely that a vision system designed for aquatic platforms could operate with aerial photos or vice versa because the pictures would be considerably different. There is a research that propose a computer vision system based on deep learning for algae monitoring. A computer vision system that allows algae monitoring to be carried out on a variety of platforms would democratize algae monitoring and make it accessible to people all over the world. The system is fast, accurate result and cheap. The researcher was installed on any robotic platforms such as USVs and UAVs for algae monitoring. This study compare object detectors based on deep learning that apply on computer vision such as Faster R-CNN, R-FCN and SSD to achieve the objective of this study. The researchers came to the conclusion that an algal monitoring system based on the R-FCN model would be extremely reliable, accurate, and quick, allowing for effective real-time algae monitoring (Samantaray et al. 2018).

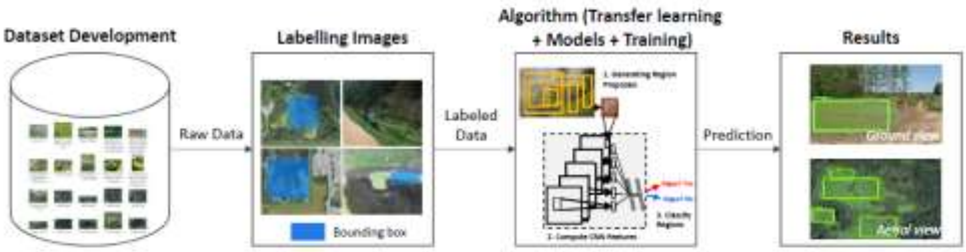

Figure 2.1: The procedure computer vision system for algae detection.

If algae is present the reliable algae monitoring system should call a high Recall, which means the system should recognize it with high Precision. To test our system's ability to recognize algae, this study performed a binary classification between two sets of photos, one with algae-containing water bodies and the other without algae-containing water bodies. The Precision values of the Faster RCNN and R-FCN were virtually identical. However, R-high FCN's Recall value shows that it is extremely robust and nearly always detects an algae bloom if one is actually present.

2.4 RESERVOIR OPERATION RULES

A studies has been done by analysis of water level fluctuations at the Xiangxi River and Three Gorges Reservoir. The Xiangxi River's water movements and algal blooms are simulated and evaluated under various scenarios. It show the quantity of chlorophyll concentration keep increasing in the water and it show grow of algal bloom. Short-term (daily operation), medium-term (about two weeks) and long-term (approximately two months) operation are the three types of TGHS water discharge processes (for the non-flood season). Daily water discharge procedure of TGHS in non-flood season is scheduled according to daily inflow, release requirement, and load demand with peak-load management for short-term operation, due to an important function of power generation and water supply to downstream river.

The amount of water discharged does not fluctuate over time, although it does change linearly during a transition period. The water discharge during the nighttime peak-load period (peak-load flow) is the same as the water discharge during the valley-load period (valley-load flow) plus the flow differential (Figure 2). Water discharges in the morning peak-load period or the medium-load period are calculated using a ratio of 0.85 and 0.65 for the morning and evening peak-load periods, respectively.

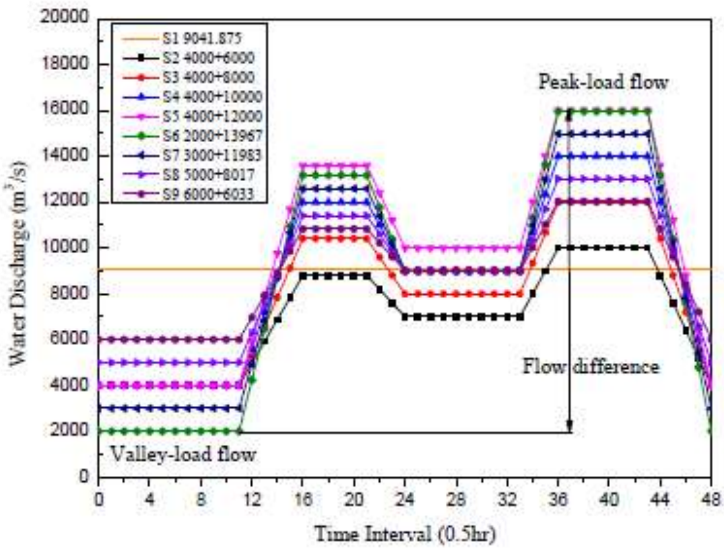

Figure 2.2: Water discharge scenarios for short-term operation.

The water level change over time is adjusted to three alternative scenarios for the medium-term operation: steady, rising, or decline (Figure 3). In addition, throughout the non-flood season, the long-term operation is planned step by step, with occasional medium-term operations and short-term operations lasting about a month. The long-term operation rules are obtained from the inference based on the above two operation rules without individual simulation.

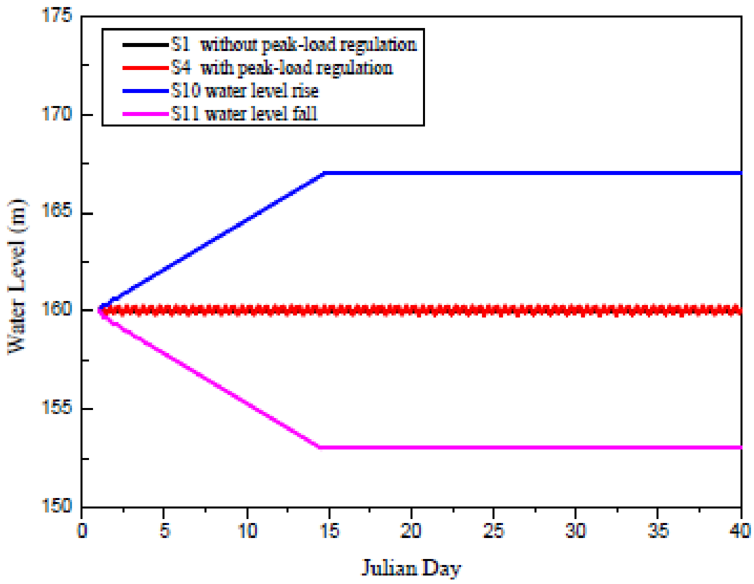

Figure 2.3: 4 types of water level scenarios for medium-term operation.

Because the 11 scenarios have the same nutritional, light, and meteorological variables, the discrepancies in simulation results on algal blooms are primarily attributable to differences in hydrodynamic circumstances owing to reservoir operating changes. Because of the water temperature difference between the two river branches, water from the mainstream plunges from the top layers into XXR and outflow of XXR discharges from the lower layers stably, according to the simulated basic flow field of XXR.

To conclude this study, Algal blooms in TGR near-dam tributaries during the non-flood season are caused by slow water movement, which creates an ideal environment for algae growth, along with unfavorable meteorological circumstances. The current TGHS scheduling rules along with scenario simulations, were used to develop reservoir operation guidelines. The reservoir operation will help to improve water quality by increasing water exchange between TGR's mainstream and the XXR tributary, resulting in more water from the mainstream plunging into XXR, resulting in increased dilution and the acceleration of algae upstream transit. For conclusion, the increasing the water level could prevent the occurrence of HAB. This was happen because the eutrophication stratum became vertically blended when more water was added and it can reducing the time for the algae remained on the surface of the water at the bay (Lian et al. 2014).

2.5 LANDSAT THEMATIC MAPPER (TM) DATA

There are a study that predicting the algae bloom by using the Landsat thematic Mapper (TM) data at the Techi reservoir, Taiwan as study site. The Landsat TM are multispectral scanning based on satellite for earth observing. To establish statistical connections to dinoflagellate algae concentrations, the researcher uses ratios of logarithm converted radiance values from Landsat Thematic Mapper (TM) data.To perform this method, numerous linear regression models has been applied. The tentative regression model required to be examined in depth for curvature and interaction effects during the algae bloom prediction regression model building procedure. Residual plots can help to decide whether one model is better than another. Following the setup of the prediction regression models, a number of residual plots and analysis were used to look for any lack of fit, outliers, or influential observations. After the residual plots and analysis were completed, this study were able to acquire the preliminary prediction regression models. Finally, we performed model validation to determine whether the prediction regression models can accurately predict the prediction destination.

The algae bloom prediction model was built using independent variables that matched the cell densities of dinoflagellate algae. In figure 2.4, it show how the process building a regression model of algal bloom prediction in this study.

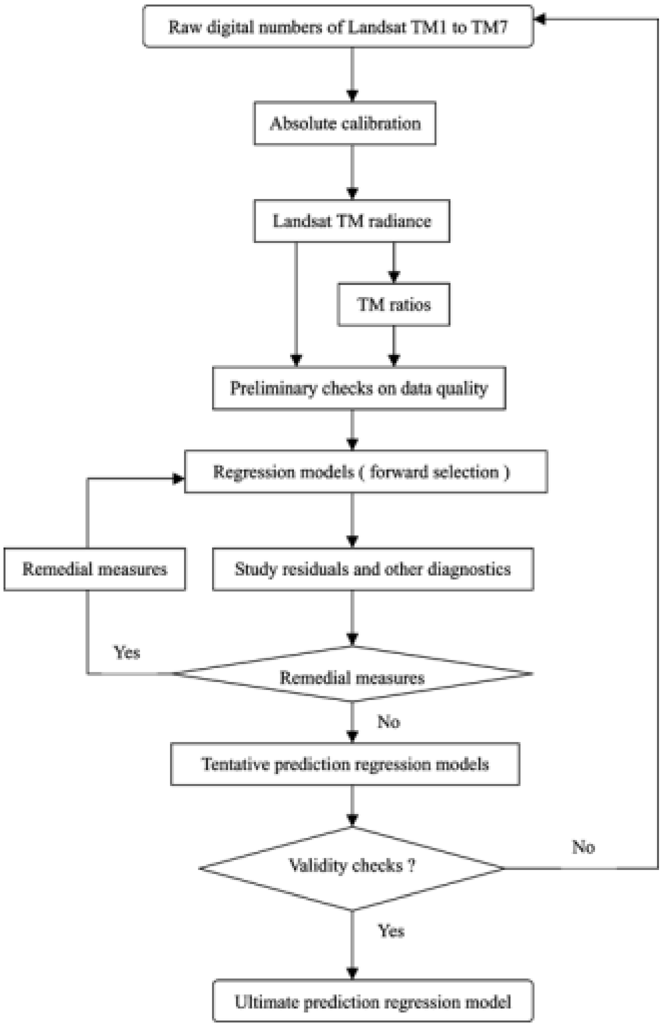

Figure 2.4: The Flow Chart Regression Model of Algal Bloom Prediction.

The result only get 74% accuracy of prediction and it because of lack spectral sensitivity and spatial resolution of the scanning device. From the poor of percentage of prediction accuracy, it still can provide a good monitoring model to observe the water quality at the reservoir and can compare with the other reservoirs at the same time(Chang et al, 2004).

2.6 Bootstrapping Based Multiple Imputation Algorithm And Binomial Logistic Regression Model

Another studies focused on weather conditions (water temperature and air temperature) rather than nutrient of the water. In this study, the researcher wanted to find the relationship between outbreak probabilities of cyanobacteria bloom each month and environmental factors such as weather conditions during the research period from 1998 to 2009. It apply the data from 1998 to 2009 of Waihai part of Dianchi Lake, China at bootstrapping based multiple imputation algorithm and for the detect probability of the outbreak this studies use the binomial logistic regression model. The figure 5 below show the analytical process for a cyanobacteria bloom.

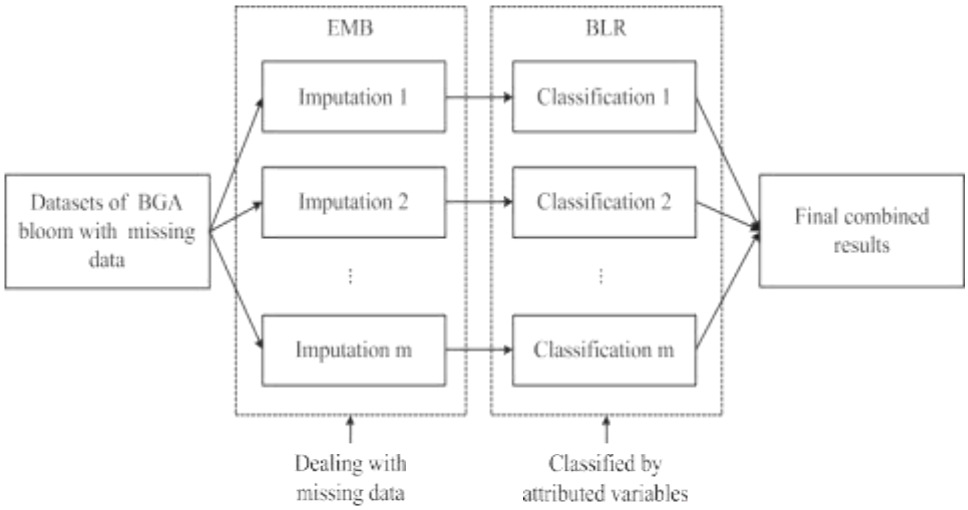

Figure 2.5: Analytical process for a cyanobacteria bloom.

From the process above, the result show the air temperature and water temperature have positive result effect to be one the factor of this algae bloom. In this study also nutrient of water were not significant factor of this outbreak (Sheng et al. 2012). From the result, more study is needed to figure out how to deal with the any elements at play and how to restore the ecosystem to prevent BGA bloom outbreaks.